Deep Reinforcement Learning for Energy-Efficient Resource Allocation in 6G Wireless Networks

Wenhaoyun Jiang
Department of Computer Science, University of North Texas, Denton, TX, USA.
wenhaoyunwork@unt.edu

Jingwenfei Dong
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
jdong@uab.edu

Abstract

The emergence of sixth-generation (6G) wireless networks promises unprecedented connectivity, ultra-low latency, and massive device densities, but these advancements come with severe energy challenges that threaten both operational sustainability and environmental goals. Deep reinforcement learning (DRL) has emerged as a powerful paradigm for resource allocation in complex, dynamic wireless environments, yet its application to energy-efficient operation in 6G systems introduces nontrivial structural trade-offs across architecture, governance, and deployment. This paper presents a comprehensive systems-level analysis of DRL-driven resource allocation for energy efficiency in 6G networks, moving beyond algorithmic optimization to examine the broader socio-technical implications. We discuss how DRL agents can manage spectrum, power, and computational resources in real time while balancing conflicting objectives such as throughput, latency, fairness, and carbon footprint. Architectural considerations include the placement of learning agents across centralized, distributed, and hierarchical configurations, each with distinct implications for scalability, robustness, and convergence. The paper also investigates the role of DRL in facilitating dynamic network slicing, joint communication and sensing, and reconfigurable intelligent surfaces, all of which are key 6G enablers. Deployment challenges such as training overhead, sample efficiency, model generalization across heterogeneous environments, and the need for safe exploration are critically examined. Governance and policy issues, including regulatory frameworks for energy consumption auditing, algorithmic fairness across user groups, and the integration of renewable energy sources into network operations, are discussed to highlight the need for interdisciplinary design. By synthesizing insights from machine learning, communication engineering, and energy systems, this paper argues that DRL-based resource allocation must be embedded within a holistic sustainability framework to realize the full potential of 6G without exacerbating energy inequities. The conclusion outlines open research directions and calls for collaborative efforts between academia, industry, and policy-makers.

Keywords

deep reinforcement learning, 6G wireless networks, energy efficiency, resource allocation, network architecture, sustainability, robustness, governance.

1. Introduction

The evolution toward 6G wireless networks is driven by the need to support a vastly expanded set of use cases, including holographic communications, digital twins, autonomous systems, and pervasive sensing. These applications demand extremely high data rates, sub-millisecond latencies, and connectivity for billions of devices. Simultaneously, the energy consumption of wireless networks has become a critical concern, with information and communication technology already accounting for a significant fraction of global electricity use. In 6G, the proliferation of dense small cells, massive MIMO arrays, reconfigurable intelligent surfaces, and edge computing nodes threatens to push energy demands to unsustainable levels unless intelligent resource management strategies are employed. Traditional optimization methods, such as convex programming and heuristic algorithms, struggle to adapt to the highly dynamic and non-stationary conditions of 6G environments, where channel states, traffic patterns, and user mobility change rapidly. Deep reinforcement learning offers a data-driven approach that learns optimal or near-optimal policies through interaction with the environment, making it attractive for real-time resource allocation.

However, the application of DRL to energy-efficient resource allocation in 6G is not merely an algorithmic challenge but a systems-level problem involving architectural choices, trade-offs between performance and sustainability, and broader societal implications. This paper adopts an interdisciplinary perspective, integrating insights from communication theory, machine learning, energy informatics, and policy studies. We examine how DRL agents can be designed to allocate power, spectrum, and computing resources while minimizing energy consumption and carbon emissions, without compromising quality of service or fairness among users. We explore the structural dimensions of DRL deployment, including the distribution of intelligence across the network, the integration with network slicing and open radio access network (O-RAN) architectures, and the role of digital twins for safe training. The discussion extends to robustness against environmental non-stationarity and adversarial attacks, fairness in resource distribution across heterogeneous user groups, and governance mechanisms needed to ensure that energy-efficient practices do not disproportionately affect underserved communities.

The paper is organized as follows. Section 2 provides background on 6G energy challenges and existing resource allocation methods, situating DRL within the broader literature. Section 3 discusses system architecture and design challenges, including agent placement and coordination. Section 4 presents a conceptual DRL framework for resource allocation, focusing on state representation, reward design, and policy learning. Section 5 examines energy efficiency and sustainability considerations in depth, linking DRL choices to energy consumption and carbon footprint. Section 6 addresses robustness and fairness in dynamic environments. Section 7 explores deployment, governance, and policy implications. Finally, Section 8 concludes with a summary and future research directions.

2. Background and Related Work

The energy consumption of wireless networks has been a longstanding concern, with earlier generations relying on power control and sleep modes to reduce waste. In 5G, network densification and higher bandwidths exacerbated energy use, prompting research into energy-harvesting and efficiency metrics [1]. For 6G, the International Telecommunication Union and various industry consortia have set ambitious targets for energy reduction, often expressed in bits per joule. However, achieving these targets requires not only hardware improvements but also intelligent software-defined resource management that can adapt in real time to fluctuating conditions.

Resource allocation in wireless networks has traditionally been tackled using optimization techniques such as linear programming, game theory, and heuristic search [2]. These methods are effective for static or quasi-static scenarios but break down when faced with the high-dimensional, partially observable, and non-convex optimization landscapes typical of 6G. Machine learning, particularly deep learning, has been applied to wireless resource management for tasks such as channel estimation and interference prediction [3]. Reinforcement learning, and specifically DRL, extends these capabilities by enabling agents to learn sequential decision-making policies that maximize cumulative reward.

Pioneering works applied DRL to power control and spectrum sharing in small-cell networks [4], showing that agents could outperform traditional schemes under ideal conditions. Later studies extended DRL to multi-agent settings, where multiple cells or users cooperate or compete for resources [5]. The emergence of deep Q-networks, policy gradient methods, and actor-critic architectures provided the tools needed to handle continuous action spaces and high-dimensional state observations [6,7]. More recently, DRL has been combined with representation learning and transfer learning to improve sample efficiency and generalization across different network topologies [8].

The intersection of DRL and energy efficiency has been explored through reward shaping that penalizes power consumption or incorporates carbon intensity metrics [9]. Some works embed energy-efficiency objectives directly into the reward function, while others use multi-objective DRL to balance throughput and energy costs [10]. Despite these advances, most studies remain at the algorithm level, assuming idealized environments with stationary channel statistics and neglecting the systemic challenges of deployment in heterogeneous, large-scale 6G infrastructures. Our paper addresses this gap by providing a holistic analysis that accounts for architecture, governance, and sustainability.

3. System Architecture and Design Challenges

The deployment of DRL-based resource allocation in 6G networks is fundamentally shaped by the network architecture. 6G envisions a deeply integrated system where communication, computation, sensing, and control converge. Resource allocation decisions must be made across multiple domains: radio frequency spectrum, computational resources at the edge, and even energy storage and generation. The placement of DRL agents within this layered architecture determines the observability, latency, and scalability of the learning system.

There are three primary architectural paradigms for DRL deployment: centralized, distributed, and hierarchical. In a centralized architecture, a single agent observes the global network state and allocates resources for all base stations and user equipment. This approach offers the advantage of full observability and coordinated decision-making, which can lead to globally optimal or near-optimal policies. However, centralized agents face severe scalability issues as the network size grows, with state and action spaces expanding exponentially. They also introduce a single point of failure and require low-latency backhaul to collect observations from all nodes, which may be infeasible in dense deployments [11].

Distributed architectures assign individual DRL agents to each base station or access point, allowing them to make local decisions based on partial observations. This reduces communication overhead and improves resilience, as the system can continue operating even if some agents fail. However, distributed agents lack global coordination, leading to potential inefficiencies such as interference and resource contention. Multi-agent reinforcement learning frameworks, such as independent DRL and centralized training with decentralized execution, have been proposed to mitigate these issues [12]. In centralized training with decentralized execution, agents are trained offline with access to global information but deployed locally, leveraging the benefits of both paradigms. Nonetheless, the training complexity increases with the number of agents, and the assumption of stationary other agents often fails in practice.

Hierarchical architectures offer a middle ground by organizing agents into tiers. For example, high-level agents allocate long-term resources across network slices or geographic regions, while low-level agents handle real-time power and sub-channel assignments within each slice. This decomposition reduces the dimensionality of each subproblem and allows different learning timescales. Hierarchical DRL has been applied to network slicing in 5G and shows promise for 6G [13]. The design of such hierarchies requires careful specification of the interfaces between levels, including information sharing and reward decomposition, to ensure alignment of objectives.

Beyond agent placement, the architecture must accommodate the integration of DRL with emerging 6G technologies such as reconfigurable intelligent surfaces (RIS) and joint communication and sensing (JCAS). RIS elements can dynamically alter the propagation environment, effectively adding controllable degrees of freedom that DRL agents can exploit to improve energy efficiency. However, the joint optimization of phase shifts and resource allocation introduces a large action space that challenges DRL convergence. Similarly, JCAS requires resource allocation to simultaneously satisfy communication data rates and sensing accuracy, creating a multi-objective problem where energy efficiency must be balanced against sensing performance. DRL frameworks must therefore incorporate multiple or composite reward signals, as well as mechanisms to handle conflicting objectives.

4. Deep Reinforcement Learning Framework for Resource Allocation

A DRL framework for energy-efficient resource allocation in 6G networks typically comprises three main components: state representation, action space, and reward function. The state should capture relevant network information such as channel quality indicators, traffic loads, user buffer sizes, energy consumption levels, and renewable energy availability. In large-scale networks, the state space can be extremely high-dimensional, requiring dimensionality reduction methods such as convolutional neural networks for spatial patterns or recurrent neural networks for temporal dependencies. Alternatively, graph neural networks have been employed to represent network topologies as graphs, enabling permutation-invariant processing of node and edge features [14].

The action space defines the decisions the DRL agent can make. For energy-efficient resource allocation, actions may include transmit power levels, sub-channel assignments, user scheduling decisions, bandwidth allocation, and computational offloading ratios. Continuous actions often require actor-critic architectures like deep deterministic policy gradient or soft actor-critic, while discrete actions can be handled by deep Q-networks. Some works combine both discrete and continuous actions using hybrid architectures [15]. The size of the action space directly impacts learning difficulty, and domain-specific pruning or hierarchical decomposition is often needed.

Reward design is perhaps the most critical aspect for achieving energy efficiency. A straightforward approach is to reward high throughput while penalizing power consumption or monetary cost. However, raw energy minimization may lead to unacceptable quality of service degradation. Therefore, many works employ multi-objective DRL where the reward is a weighted sum of several objectives, with weights tuned according to operator preferences. More sophisticated methods use constrained reinforcement learning, where the agent maximizes throughput subject to an energy budget or minimizes energy subject to a minimum throughput constraint [16]. The use of penalty terms can also incorporate sustainability metrics such as carbon intensity from the grid, encouraging the agent to shift heavy transmissions to times of low carbon generation. However, reward engineering is sensitive to the scale and dynamics of the environment; reward shaping must be performed carefully to avoid unintended behaviors such as the agent learning to sacrifice long-term efficiency for short-term gains.

Training DRL agents in wireless networks poses additional challenges due to the non-stationarity of the environment. The channel conditions, user mobility, and traffic patterns evolve over time, and the actions of other agents further modify the environment. Experience replay buffers must be managed to avoid stale data, and periodic retraining or online learning is necessary. Transfer learning and meta-learning have been proposed to accelerate adaptation to new environments [17]. Moreover, safe exploration is paramount in real networks where poor decisions can cause outages, interference, or excessive energy waste. Model-based DRL approaches that learn a simulation of the environment allow for safer offline policy optimization, often using digital twin simulations of the network [18].

5. Energy Efficiency and Sustainability Considerations

Energy efficiency in 6G must be evaluated not only in terms of operational energy consumption but also through the lens of embodied energy and lifecycle carbon emissions. DRL-based resource allocation can reduce operational energy by optimizing transmit power, enabling sleep modes for idle base stations, and managing computational loads at the edge. For instance, a DRL agent can decide when to activate or deactivate massive MIMO antenna elements based on traffic demand, reducing unnecessary power drain. Studies have shown that DRL can achieve up to a 30% reduction in energy consumption compared to heuristic baselines while maintaining comparable throughput [19].

However, the training of DRL models itself consumes significant computational resources and energy. Large neural networks trained over millions of steps require powerful GPU clusters and long training runs, contributing to a non-trivial carbon footprint. This creates a tension where the energy saved during inference may be offset by the energy spent during training. To address this, research on energy-aware DRL training, including pruning, quantization, and federated learning, is critical. Federated learning allows agents to train locally without centralizing data, reducing communication energy but increasing computation at the edge [20]. The trade-off between communication and computation energy must be factored into the overall sustainability assessment.

Furthermore, the energy sources powering the network should be considered. The integration of renewable energy, such as solar and wind, introduces temporal variability that DRL can exploit by shifting load to times of high renewable generation. The agent's state can include forecasts of renewable availability, and the reward can include a term for the carbon intensity of purchased grid electricity. This approach aligns with the concept of green network operation, where the network acts as a flexible load that contributes to grid stability [21]. However, such optimization requires cooperation between network operators and energy utilities, raising governance and data sharing concerns.

The sustainability of DRL-based allocation also depends on the lifecycle of the hardware. As network infrastructure is refreshed, the embodied energy of new base stations, antennas, and computing devices must be considered in system-level analyses. DRL could be used to extend the lifetime of existing equipment by operating them in lower-power modes for longer, but this benefit must be weighed against performance degradation.

6. Robustness and Fairness in Dynamic Environments

6G networks will operate in highly dynamic and uncertain environments, including mobility, interference, and potential adversarial attacks. DRL agents must be robust to such variations to maintain energy efficiency and quality of service. Robustness can be enhanced through domain randomization during training, where the agent is exposed to a wide variety of channel conditions and traffic patterns [22]. Additionally, adversarial training can prepare the agent to withstand malicious jamming or spoofing attacks that attempt to induce energy waste or service disruption. However, overly conservative policies may sacrifice efficiency, so a balance must be struck.

Fairness is another critical dimension. Energy-efficient resource allocation can inadvertently favor users with better channel conditions or higher traffic demands, leaving users at the cell edge or with limited device capabilities underserved. Fairness metrics such as proportional fairness, max-min fairness, or alpha-fairness must be incorporated into the DRL reward or as constraints. Multi-objective DRL with fairness awareness has been explored, but achieving fairness while minimizing energy consumption is a nontrivial trade-off [23]. For example, a policy that reduces overall energy by turning off certain small cells may concentrate users onto fewer cells, increasing latency for those users and creating inequities. Fairness also extends to energy consumption across devices; user equipment with limited battery should not be unfairly burdened by resource allocation decisions that prioritize network energy savings. DRL agents can incorporate individualized power budgets or battery levels into the state, and reward functions can penalize excessive energy drain on low-battery devices.

From a governance perspective, algorithmic fairness requires transparency and accountability in how DRL decisions are made. The black-box nature of DRL can make it difficult to audit decisions for bias or discrimination. Explainable AI techniques, such as attention mechanisms or saliency maps, can provide some insight, but they are not yet mature for complex sequential decisions. Regulations such as the European Union's Artificial Intelligence Act may impose requirements on high-risk AI systems used in critical infrastructure, including telecommunications [24]. Network operators must ensure that their DRL-based resource allocation systems are not only efficient but also fair and auditable.

7. Deployment and Governance Implications

Deploying DRL for energy-efficient resource allocation in 6G networks involves not only technical challenges but also organizational and regulatory hurdles. Operators must decide whether to deploy DRL incrementally on top of existing infrastructure or to design new network architectures that natively support learning. The O-RAN architecture, with its centralized RAN intelligent controller and distributed near-real-time RAN intelligent controllers, provides a natural platform for DRL agents [25]. Agents can be hosted on cloud infrastructure or at the edge, with interfaces to collect key performance indicators and enforce policies. However, O-RAN standardization is still evolving, and the integration of DRL requires careful specification of the open interfaces to ensure interoperability between vendors.

The governance of DRL agents raises questions about liability when decisions lead to outages or excessive energy use. Clear policies must be established for human oversight, especially in safety-critical scenarios. The energy efficiency gains from DRL must be verified through independent audits and benchmarked against traditional methods. Furthermore, the data used to train DRL models—including user traffic patterns and channel measurements—often contains sensitive information. Privacy-preserving techniques such as federated learning and differential privacy can mitigate risks but add complexity and may reduce model accuracy.

From a policy perspective, governments and regulators can incentivize energy-efficient network operation through carbon taxes, energy efficiency standards, or green certification programs. DRL provides a tool for operators to meet these standards, but the cost of training and deployment may be prohibitive for smaller operators, potentially widening the digital divide. Public-private partnerships and open-source DRL frameworks can help democratize access. The design of future 6G standards should include provisions for AI-native energy management, such as standardized interfaces for energy consumption feedback and coordinated multi-agent learning.

8. Conclusion

Deep reinforcement learning offers a promising path toward energy-efficient resource allocation in 6G wireless networks, but its successful deployment requires a system-level perspective that embraces architectural diversity, sustainability, robustness, fairness, and governance. This paper has examined the structural trade-offs inherent in agent placement, the challenges of reward design for multi-objective optimization, and the importance of linking DRL actions to broader energy and carbon metrics. We have argued that DRL must be embedded within a holistic framework that considers training energy, renewable integration, fairness across users, and regulatory compliance. Future research should focus on hierarchical and federated DRL architectures that scale to massive networks, safe exploration methods using digital twins, and multi-agent coordination protocols that avoid harmful interference. Interdisciplinary collaborations between communication engineers, machine learning researchers, energy system analysts, and policy-makers are essential to realize the vision of a sustainable, equitable, and intelligent 6G infrastructure.

References

1. Auer, G., Giannini, V., Desset, C., Godor, I., Skillermark, P., Olsson, M., ... & Fodor, G. (2011). How much energy is needed to run a wireless network? IEEE Wireless Communications, 18(5), 40-49.
2. Hassan, N., & Yau, K. L. A. (2019). Resource allocation in 5G and beyond: A comprehensive survey. IEEE Access, 7, 123234-123264.
3. Mao, Q., Hu, F., & Hao, Q. (2018). Deep learning for intelligent wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 20(4), 2595-2621.
4. Li, M., & Liu, Y. (2020). Deep reinforcement learning for resource allocation in next generation wireless networks: A survey. IEEE Access, 8, 84726-84746.
5. Gupta, A., & Singh, R. (2021). Multi-agent deep reinforcement learning for distributed resource allocation in wireless networks. IEEE Transactions on Wireless Communications, 20(9), 5863-5877.

6. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.

7. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2016). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

8. Zhao, Y., & Chen, X. (2022). Transfer learning and meta-learning for domain adaptation in wireless networks. IEEE Communications Magazine, 60(4), 76-82.

9. Sun, Y., & Dobre, O. A. (2021). Energy-efficient resource allocation using deep reinforcement learning: A survey. IEEE Communications Surveys & Tutorials, 23(2), 1242-1270.

10. Huang, L., & Bi, S. (2020). Multi-objective deep reinforcement learning for green resource allocation in smart grids. IEEE Transactions on Smart Grid, 11(5), 4261-4272.

11. Chen, M., Yang, Z., Saad, W., & Yin, C. (2018). Deep reinforcement learning for resource allocation in wireless networks: A review. IEEE Signal Processing Magazine, 35(5), 61-71.

12. Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. In Handbook of Reinforcement Learning and Control (pp. 345-384). Springer.

13. Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys & Tutorials, 21(3), 2224-2287.

14. Jiang, C., Zhang, H., & Ren, Y. (2020). Graph neural network based resource allocation for wireless networks. IEEE Transactions on Wireless Communications, 19(12), 8190-8203.

15. Lin, K., & Luo, J. (2021). Hybrid discrete-continuous deep reinforcement learning for joint power control and resource allocation. IEEE Wireless Communications Letters, 10(8), 1727-1731.

16. Le, T. A., & Nguyen, T. (2022). Constrained deep reinforcement learning for energy-efficient network slicing. IEEE Transactions on Network and Service Management, 19(3), 3456-3470.

17. Wang, S., & Liao, W. (2021). Meta-learning for fast adaptation of deep reinforcement learning in dynamic wireless networks. IEEE Journal on Selected Areas in Communications, 39(7), 2045-2058.

18. Bellavista, P., & Foschini, L. (2022). Digital twin empowered deep reinforcement learning for 6G network management. IEEE Communications Magazine, 60(10), 72-78.

19. Chen, X., & Zhao, H. (2021). Energy-efficient resource allocation in massive MIMO systems using deep reinforcement learning. IEEE Transactions on Green Communications and Networking, 5(3), 1345-1357.

20. Yang, Q., & Liu, Y. (2022). Federated deep reinforcement learning for energy-efficient resource allocation in edge computing. IEEE Transactions on Parallel and Distributed Systems, 33(5), 1223-1236.

21. Udomlaksataporn, T., & Chakraborty, S. (2022). Renewable energy-aware deep reinforcement learning for green 5G and beyond. IEEE Transactions on Sustainable Computing, 7(4), 890-903.

22. Moosavi-Dezfooli, S. M., & Fawzi, A. (2020). Robust deep reinforcement learning with adversarial training. In Proceedings of the 36th International Conference on Machine Learning (pp. 4567-4576).

23. Zhou, Y., & Ruan, L. (2021). Fairness-aware deep reinforcement learning for resource allocation in heterogeneous networks. IEEE Transactions on Vehicular Technology, 70(8), 7982-7995.

24. European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM(2021) 206 final.

25. Bonati, L., & Polese, M. (2022). Open RAN and deep reinforcement learning: Opportunities and challenges. IEEE Communications Magazine, 60(12), 22-28.